Analogical Modeling: Exemplars, Rules, and Quantum Computing

ROYAL SKOUSEN
Brigham Young University

1. Properties of analogical modeling

I begin this paper by outlining the main properties of analogical modeling (AM):

- AM is an exemplar or instance-based system of prediction; it is not a rule-based system nor a neural network.

- AM is a procedural system, not a declarative one; one cannot directly find its predictions within the system itself. Instead, predictions are always made in terms of a given context for which a prediction (or outcome) is sought.

- AM is not a nearest neighbor approach; it does include nearest neighbors in its predictions, but it also regularly uses non-neighbors to make predictions. However, such non-neighbors can be used only under a well-defined, explicit condition of homogeneity.

- No training stage occurs in AM, except in the trivial sense that one must collect exemplars in order to make predictions. There is no setting of parameters nor any prior determination of variable significance. The significance of any combination of variables is always determined in terms of the given context for which we seek a predicted outcome. Each prediction is done "on the fly".

- AM has a simple measure of uncertainty that is defined by the number of outcome disagreements between pairs of exemplars.

- AM uses a simple decision principle to determine homogeneity – namely, minimize the number of disagreements. No increase in uncertainty is permitted, which means that no analogical analysis will allow any loss of information.

- This definition of homogeneity results in two types of homogeneous spaces in an analogical analysis: (1) the subspace is *deterministic* in behavior, which means that there will be no disagreements at all within that contextual space and therefore no a priori limitation on the extent of a deterministic subspace; (2) the subspace is *non-deterministic* in behavior, which means that disagreement occurs within the contextual space, but is restricted to one subspace of the entire space.

- The resulting probability of using a particular exemplar depends upon three

factors: (1) *proximity:* the closer the exemplar to the given context, the greater its chances of being selected as the analogical model; (2) *gang effect:* when a group of exemplars in the same space behave alike, the chances of one of those exemplars being selected is multiplied; (3) *heterogeneity:* the chances of an exemplar being used is zero whenever there is any intervening exemplar closer to the given context that behaves differently.

We can see these three properties in the following schematic. Here we have a class of possible exemplars for analogically predicting the English plural of the nonce noun *ux*. For purposes of illustrating the properties of analogical modeling, the dataset for predicting the given context (*ux*) is arbitrarily restricted to nouns ending in the spelling *x*. There are 23 exemplars in the dataset, of which most take the regular plural, *-es*, but two have exceptional plurals: (1) the nearest-neighbor *ox*, which normally takes the irregular ending *-en* (thus, *ox/oxen*) but infrequently is regularized (*ox/oxes*), and (2) the distant *index*, which has two plurals, each about equally probable: the regular *indexes* and the Latinate *indices*.

Analogically Predicting the Plural for *Ux*
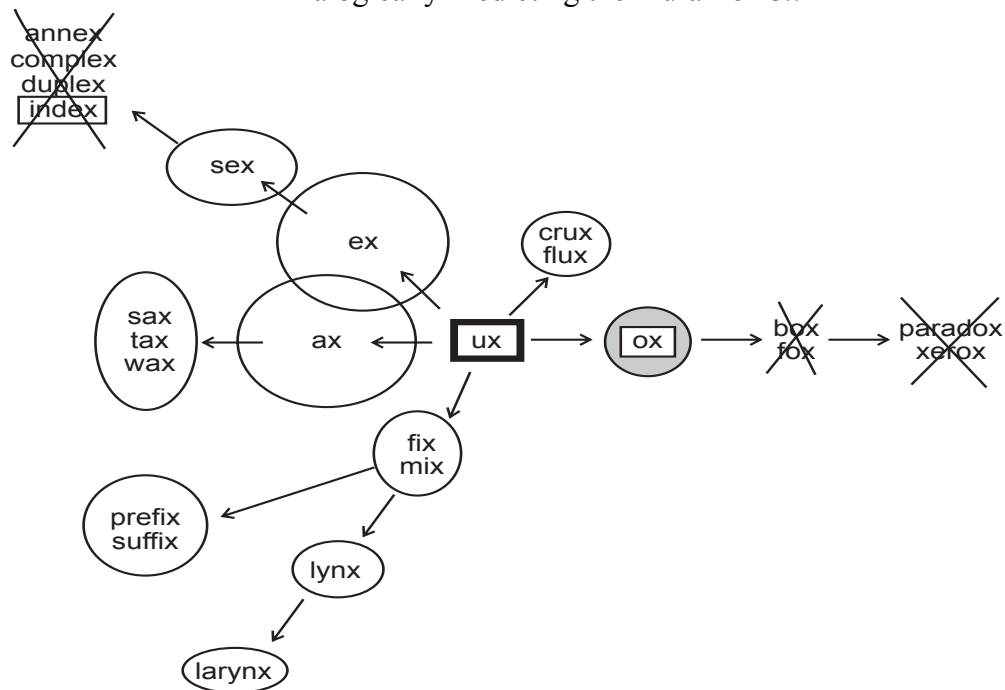
In this schematic, the given context (*ux*) is in a heavy square box. The given context is surrounded by its nearest neighbors, *ax, ex, crux, flux,* and *ox*. The size of the circles surrounding each group of neighbors approximately shows the chances of selecting each of these exemplars as the analogical model for predicting the plural for

ux. The circles for *ax* and *ex* are the largest because they are supported by various gangs of less-near neighbors that behave like *ax* and *ex* (namely *sax, tax, wax, sex, fix, mix, lynx, larynx, prefix,* and *suffix*). The further we go out from the given context, though, the smaller the circles become, thus showing a decrease in choosing less-near neighbors because of their relative lack of proximity to *ux.* Although *crux* and *flux* are nearest neighbors, their strength is less because they have no support from less-near neighbors.

In contrast, the behavior of *ox* leads to heterogeneity in the contextual space. There is some chance that *ox,* a nearest neighbor, will serve as the analogical model for *ux*, which would predict the plural *uxen.* But there are no other exemplars that will support *ox* in predicting *uxen.* Even regularly behaving exemplars like *box, fox, paradox,* and *xerox* will have no influence on predicting the plural for *ux* since the differently behaving *ox* is closer to *ux* and stands between the given context and those less-near exemplars. For this reason, these four outliers, even though they are regular in behavior, occur in a heterogeneous subspace of the contextual space and are therefore totally excluded from the analogical set. Of course, there are enough regular exemplars still in the analogical set to allow the regular behavior to be overwhelmingly predicted. But because of the proximity of *ox* to the given *ux*, there remains a small, but noticeable, possibility that the predicted plural will be *uxen.*

Finally, we note that there is a group of distant exemplars that are also excluded from the analogical set for *ux* – namely *annex, complex, duplex,* and *index*. The non-deterministic behavior of *index* (with its varying plural *indexes* versus *indices*) means that the subspace containing all four of these words is heterogeneous. Again the reason is that there are intervening exemplars closer to the given *ux* that behave differently (namely, *ex* and *sex* behave differently from *index*, at least for exemplars taking *indices* as the plural). And because *index* is therefore excluded, all the other exemplars in the same subspace (namely *annex, complex,* and *duplex*) are also eliminated from the analogical set, even though their own behavior is as regular as the intervening exemplars *ex* and *sex*. Thus we have a second example of some regularly behaving exemplars being excluded because of heterogeneity.

2. The basic literature

Analogical modeling is treated in three fundamental books: (1) *Analogy and Structure* (Skousen 1992), (2) *Analogical Modeling of Language* (Skousen 1989), and (3) *Analogical Modeling: An exemplar-based approach to language* (Skousen, Lonsdale, and Parkinson 2002).

Written in the early 1980s and finished in 1984, *Analogy and Structure* describes the fundamental properties of both rules and analogy and provides the mathematical basis for the theory of analogical modeling. This book was actually published later than *Analogical Modeling of Language,* which was written in the mid 1980s to
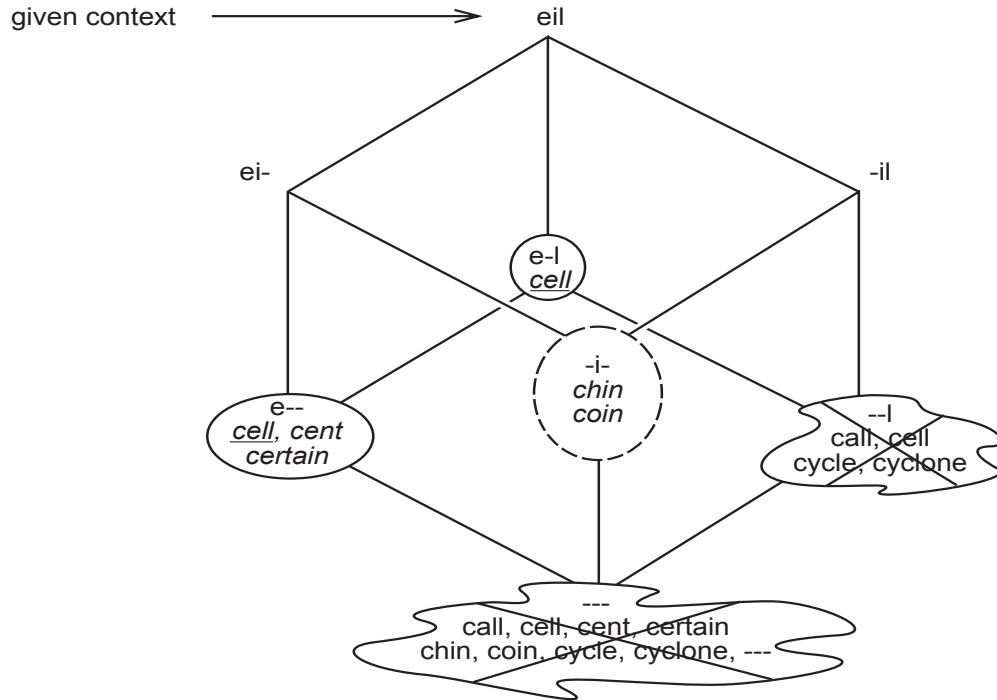
provide a more general outline of the theory of analogical modeling and, in particular, its application to language. The third book, *Analogical Modeling: An exemplar-based approach to language*, describes current developments in the theory. Besides providing a tutorial on AM and how to run the computer program, it describes the psycholinguistic evidence for AM and applies the theory to a number of specific language problems. The theory is also compared to nearest neighbor approaches. Quantum computing is proposed as the natural way to handle the computational exponential explosion that is inherent in analogical modeling.

3. The terminology of analogical modeling

In this section I provide a sample analogical set in order to explain the terminology of AM. In the following example, I use AM to predict the pronunciation of the initial *c* letter in *ceiling*. Our DATASET will be a highly restricted one consisting of 34 words with an initial *c* letter. (The words in this simple dataset can be found in Skousen, Lonsdale, and Parkinson 2002:13-14.) In this dataset, there are three possible OUT-COMES: *k, s,* and *č* (based on words like *call, cent,* and *chin*). For this particular simplified example, we will try to predict the pronunciation of the initial *c* in *ceiling* by using the first three following letters (after the initial *c*) as VARIABLES. We make our predictions in terms of a GIVEN CONTEXT, which for *ceiling* will be *e i l* . The remaining letters *i n g* will be ignored in this analysis. With three variables, there are eight possible combinations of variables that can used to predict the outcome – namely *e i l, e i −, e −l, −i l, e −−, −i −, −−l,* and *−−−*. Whenever the dash (−) is used, it means that the actual particular value of that variable is being ignored, or in other words, that variable can be assigned anything. These combinations of variables are called SUPRACONTEXTS. In general, if there are *n* variables, there will be 2^n supracontexts.

The 2^n supracontexts form a partial ordering based on subcontextual relationships – namely, we connect more general supracontexts to more specific supracontexts, providing the more specific ones are subcontexts of the more general ones. And for each supracontext we list the dataset occurrences found in that supracontext (see the diagram on the next page). Three of the supracontexts, including the given context *e i l* , have no occurrences and are thus empty. Three of the supracontexts, all encircled, are HOMOGENEOUS in behavior. At least one of two conditions must be met for a supracontext to be homogeneous: either there is only one outcome type in the supracontext or all the occurrences are found in the same subcontextual subspace. In the first case, the supracontext is deterministic in behavior since only one outcome is found. Two of the homogeneous supracontexts are deterministic – namely *e −l* and *e −−*. The occurrences in these two supracontexts take only the *s* outcome. I encircle these two deterministic supracontexts with a solid line.

given context ⟶ eil

ei-

-il

e-l
cell

-i-
chin
coin

e--
cell, cent
certain

--l
call, cell
cycle, cyclone

---
call, cell, cent, certain
chin, coin, cycle, cyclone, ---

The other homogeneous supracontext $-i-$ is non-deterministic in behavior. There are two different outcomes in this supracontext (namely *č* for *chin* and *k* for *coin*). But these two occurrences are found only within this supracontext and not in any more specific supracontext higher up in the lattice. In the above diagram, this non-deterministic homogeneous supracontext is encircled with a dashed line. Basically, a non-deterministic homogeneous supracontext permits no other occurrences between itself and the (non-occurring) given context. A deterministic homogeneous supracontext, on the other hand, does allow such intervening occurrences. For instance, the more general supracontext $e--$ contains three occurrences, of which one, *cell,* is also found in $e-l$ and thus closer to the given context than the two other occurrences, *cent* and *certain*. But all three of these have the same outcome *s,* thus homogeneity is maintained. It should also be noted here that *cell* is found in two different homogeneous supracontexts, namely $e-l$ and $e--$. In a sense, we can think of *cell* as multiply occurring. This property of multiple existence has important consequences when we compare analogical modeling to quantum mechanics.

Finally, two of the supracontexts turn out to be HETEROGENEOUS, namely $--l$ and the most general supracontext $---$. Both are non-deterministic and their occurrences are not restricted to a single subcontext. For instance, $--l$ contains *call* (with outcome *k*) and *cell, cycle,* and *cyclone* (with outcome *s*), which makes this supracontext non-deterministic. But in addition, *cell* is found in $e-l$, which is closer to the given context and thus there is an intervening different behavior. In a

similar way, the most general supracontext is also heterogeneous. In fact, it is easy to show that if a supracontext is heterogenous, a more general one containing that one as a subcontext will also be heterogeneous. We refer to this property as INCLUSIVE HETEROGENEITY. In other words, $---$ is inclusively heterogeneous because the more specific supracontext $--l$ is heterogeneous. In the above lattice, we enclose each heterogeneous supracontext with a jagged line and also place a large X through each one, thereby indicating that none of its occurrences can be accessed.

Occurrences that are far enough away from the given context are eliminated as possible exemplars. Closer occurrences can be found in homogeneous supracontexts, and those occurrences can be used as exemplars. Note, in particular, that the same occurrence can be inaccessible from a heterogeneous supracontext, but accessible from a homogeneous supracontext. Thus *cell* is accessible from both $e-l$ and $e--$ (which are homogeneous), but not from $--l$ or $---$ (which are heterogeneous).

We now use the resulting analogical set to predict the outcome. Instead of directly selecting an occurrence in a homogeneous supracontext, we rely on directional pointers that connect every pair of occurrences in a homogeneous supracontext. The basic RULE OF USAGE for predicting behavior is to randomly select one of these pointers and to observe which occurrence it is pointing to. This occurrence becomes the analogical model and the outcome associated with this exemplar is selected as the predicted outcome. We refer to this rule of usage as RANDOM SELECTION. By selecting pointers, the probability of choosing a particular homogeneous supracontext is the square of the frequency of that supracontext. Underlyingly, the frequency of the supracontext is a linear measure and is equal to the number of occurrences in the supracontext, but the number of pointers is equal to the frequency squared. Another possible rule of usage is SELECTION BY PLURALITY, which allows us to examine the occurrences predicted by random selection and choose the one that is most frequent. This rule of usage leads to the best possible decision whenever we are trying to maximize gain or minimize loss.

The analogical set can now be used to predict the pronunciation of the initial *c* in the word *ceiling*. For each homogeneous supracontext, we list the number of occurrences for each outcome, plus the number of pointers to each outcome. (See the chart on the next page.) If the underlying frequency (or number of occurrences) for a homogeneous supracontext is m, then there will be m^2 pointers, proportionally distributed. Empty supracontexts will have no occurrences or pointers and are trivially homogeneous. Every heterogeneous supracontext (marked by an *x*) may have occurrences, but none of their pointers will be accessible. For random selection, the probability of selecting the *s* outcome dominates, even though we only had 34 sample spellings in our restricted dataset. Using selection by plurality, the *s* outcome would be chosen every time.

supracontexts	*occurrences*			*pointers*		
	k	*s*	*č*	*k*	*s*	*č*
e i l	-	-	-	-	-	-
e i –	-	-	-	-	-	-
e – l	-	1	-	-	1	-
– i l	-	-	-	-	-	-
e – –	-	3	-	-	9	-
– l –	1	-	1	2	-	2
x – – l	1	3	-	-	-	-
x – – –	21	9	3	-	-	-
	totals			2	10	2
rule of usage	random selection			1/7	5/7	1/7
	selection by plurality			0	1	0

4. Applying analogical modeling to morphology

I begin this section by analyzing what is considered a classic categorical "rule" – namely the rule for predicting the indefinite article *a/an* in English. As is well-known, the crucial variable is whether the following segment is a consonant or a vowel, thus *a boy* versus *an apple.*

When we treat this problem analogically, we are able to derive the typical rule-like properties that we expect, but other significant behaviorial properties can be derived that no rule approach is able to predict or model. In the following example, the dataset for predicting *a* versus *an* is derived from two book chapters that I wrote over a year ago (and with no intention of using to construct this dataset). In all, there are 251 examples of the indefinite article in this dataset, of which 40 take *an* and 211 take *a*. Each *an* example is followed by a vowel-initial word, each *a* example by a consonant-initial word. So the data is "clean" – the two chapters were correctly edited so that there are no exceptions to the "rule".

Our dataset is described by 9 variables based on the surrounding sounds, both preceding and following – in other words, we include variables that we know in advance are not supposed to be "significant" or "crucial". We basically use the two preceding and the two following phonemes, along with their syllabic nature, and specify whether there is a phrasal break just before the indefinite article.

Using this dataset to predict the indefinite article for 8 different test items (not in the dataset), we get the following basic results for the full 251 items: (1) if the following word in the test item begins with a consonant, we get the "correct" *a* form virtually 100% of the time; (2) if the following word in the test item begins with a

vowel, we get the "correct" *an* form about 98% of the time, but for 2% of the time we get leakage towards the "incorrect" *a*. This difference in leakage is quite dramatic when we consider a restricted dataset, starting out with only a few occurrences in the dataset and steadily increasing the number of occurrences. Initially, *a* is uniformly predicted no matter what kind of segment (consonant or vowel) follows, but as the data increases and tokens of *an* are introduced, the model begins to irregularly predict *an* when followed by a vowel, but over time moves inevitably towards systematically predicting *an*. Yet the model continues to predict some leakage towards *a* (but none towards *an* when followed by a consonant), even when adult-like behavior has been achieved (after about 80 occurrences of *a* and *an*). Below we have a graph of the predicted *an* behavior for the four vowel examples, with their one-way leakage decreasing erratically until stability is reached (but not with complete predictability). Note the small window of leakage that never fully closes.
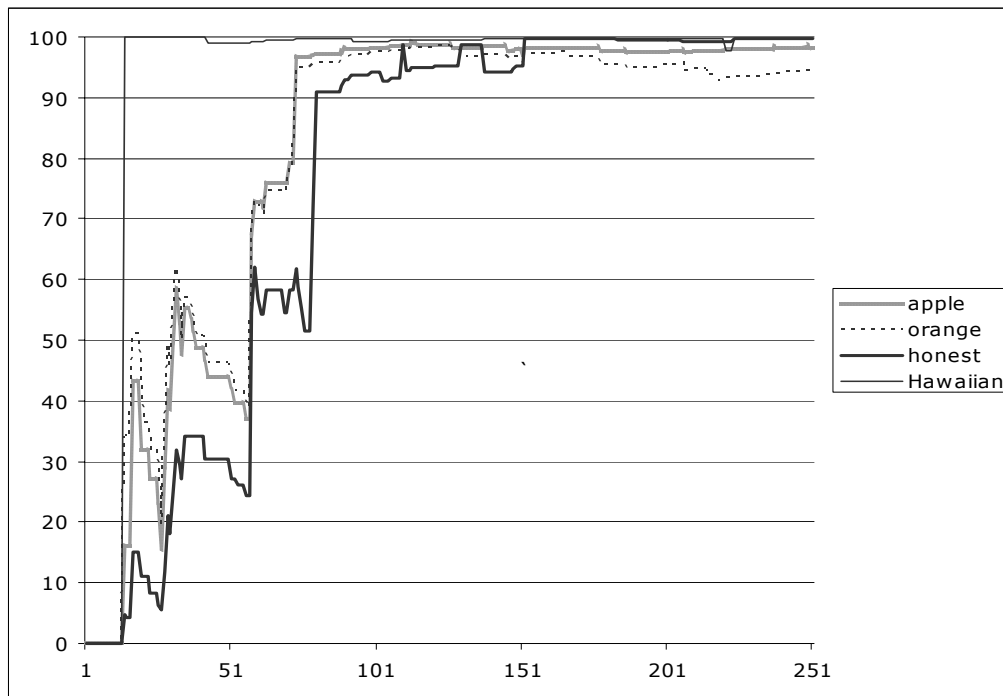

On the other hand, the expected *a* is consistently predicted when the following segment is a consonant, even when the number of occurrences in the dataset is small.

We note from this example that AM is not simply reproducing the data. The data itself is fully "regular", having no exceptions to the "rule". The predicted fuzziness of AM justifies Sapir's well-expressed statement that "all grammars leak". And sometimes the leak is directional. Unlike rule approaches, there is no need to hunt

for extra reasons to explain the performance. The system itself predicts the kinds of errors that occur, both for children learning the system and for adults who occasionally replace *an* by *a*. There is no need to set up a theory of markedness that would favor open rather than closed syllables (thus "explaining" why *an* tends to be replaced by *a,* but not vice versa). A variant explanation would be to declare that it is phonetically more difficult to pronounce *an boy* than *a boy*. This may be true enough for the youngest of children learning English, but it is not for adult speakers (or even moderately young children) who readily pronounce *one boy* as /wən bɔi/ rather than as /wə bɔi/.

The one-way leakage that we observe is inherent within the dataset itself and is due to the relative sparsity of *an* within the contextual space (when compared with *a*). This tendency can also be directly seen in the historical dynamics of the indefinite article. Using the same dataset for *a*/*an* as before (with 40 occurrences of *an* out of 251), we can model the historical drift in dialects of English towards replacing *an* completely by *a*. This particular model predicts that when there are less than about 70 accessible occurrences to predict from, there will be a steady S-transition from the original state of 40 occurrences of *an* to eventually none. Overall, the S-transition is very regular and smooth: the shifting starts out slowly with a few innovative cases, then the transition moves more rapidly, and in the end slows down, with a few relics holding out until the transition is complete. But when we consider the probabilistic predictions for each of the individual 40 cases of *an*, we see that the drift is highly erratic, with sharp shifts and frequent reversals in the predicted behavior, but eventually we get a complete shift from *an* to *a*.
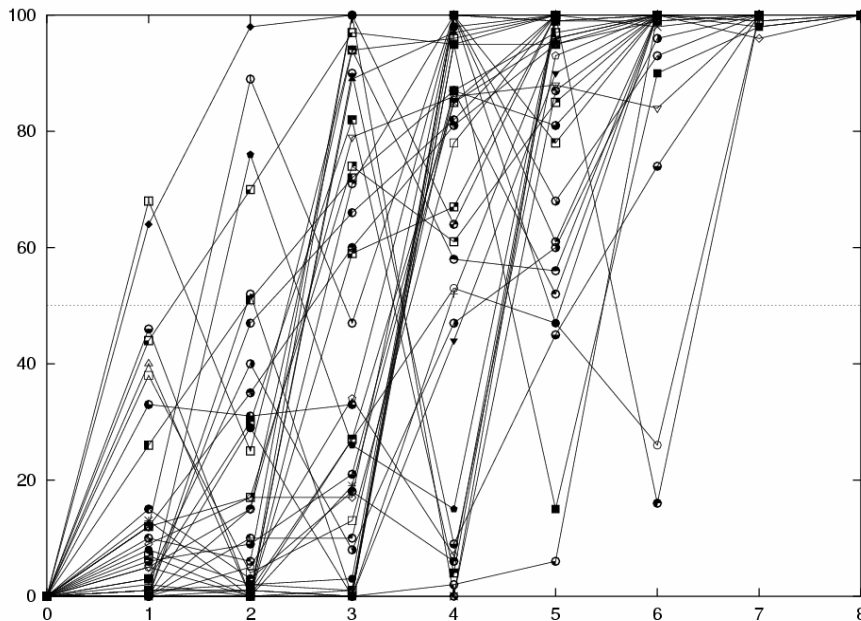
Although the overall shifting seems quite tranquil, the specific examples of the drift are very turbulent. Actual historical examples of drift reflect this kind of turbulence.

5. Robustness

One important advantage of analogical modeling is that it can deal with what might be considered unexpected or defective language data such as dialectal developments and various errors found in adults' and children's language. One particular aspect of this robustness is the ability to make predictions even when the "crucial" variables are missing. For instance, in the case of the indefinite article *a*/*an*, suppose that the first segment in the immediately following word is overlaid with noise, yet we still wish to predict whether the article should be *a* or *an*. Using AM, we are still able to make predictions when the crucial variable is missing. Sometimes the resulting analogical set is dominated by a particular word, which basically means that the system has figured out what the partially obscured word is. Other times, no specific word may dominate the analogical set, but nonetheless redundancies in the dataset allow the correct prediction to be made. For instance, if the second segment in the following word is an *s*, then the analogical set will be dominated by examples for which the first segment (here obscured in the given context) is a vowel. This result derives from the fact that there are no consonant-*s* word-initial sequences in English; only vowel-*s* sequences are found. (Greek borrowings such as *psychology* are of course not pronounced with an initial /ps/ cluster in English.)

One important property of AM is its ability to predict leakage when given contexts are near exceptions. Sometimes unusual pronunciations of certain frequent words can affect nearby words. For many speakers of English (such as myself), the names for the days of the week have an alternative /i/ vowel pronunciation for the final *-day* morpheme. For instance, *Monday* has the pronunciation /məndi/ in addition to the standard /məndei/. This alternation leads speakers to accidentally produce the /ei/ pronunciation when /i/ is correct, but only in words that are close to the names of the days of the week, such as the burger establishment *Wendy's* being pronounced (by me) as /wɛndeiz/ or the example from a friend who first pronounced the name of the Utah community *Sandy* as /sændei/, then immediately corrected it to /sændi/. The nearby words *Wednesday* and *Sunday* serve as the source for the alternative /ei/ pronunciations, but only because there is the variant /i/ that allows this kind of phonological backformation.

Sometimes exceptional behavior is created because an item is extremely close to a single exemplar. Consider the word *consonantal,* which linguists typically pronounce as /kɑnsənɛntəl/ rather than the dictionary pronunciation /kɑnsənæntəl/. The obvious nearby analogy here is the word *continental*.

Other times a strong attraction occurs because of a whole gang of exemplars. Consider the pronunciation of the word *nuclear* as /nukjələ-/ rather than the standard /nuklɪə-/. As Jesse Sheidlower, an editor for the Oxford English Dictionary has pointed out (*New York Times*, 13 October 2002), the reason for the much-maligned innovative pronunciation is a very large gang of words that end in -/kjələ-/, namely ***particular, spectacular, circular, molecular, secular, perpendicular, muscular,*** *vernacular, jocular, vascular, curricular, ocular, corpuscular, follicular, specular, testicular, vehicular,* and *vesicular*. And many of these words are fairly frequent (the ones given in bold). On the other hand, there is only one word that supports the standard pronunciation for *nuclear* – namely *cochlear,* itself quite rare. Thus it is not surprising that speakers are tending to pronounce *nuclear* as /nukjələ-/, despite its spelling.

6. Not just nearest neighbors
Sometimes it has been argued that analogical modeling could be more efficient if it only looked for the nearest neighbor(s) to select an exemplar rather than considering exemplars that might be more distant. The following case involving the Finnish past tense clearly requires exemplars that are not the nearest neighbor. In trying to predict the past tense for the Finnish verb *sorta-* 'to oppress', there are two viable possibilities: *sorsi* and *sorti*. The first can be derived from the stem *sorta-* by replacing the *t* by an *s* and the final *a* vowel by *i*. We represent this outcome by the representation *tV-si*, where *V* stands for a vowel. The second possibility, *sorti*, can be derived by simply replacing the final *a* vowel by *i* and thus leaving the *t* of the stem unchanged. We represent this outcome as *V-i*.

Rule approaches to the Finnish language have predicted that the past tense for *sorta-* should be *sorsi*, which is the historical form. Yet actually *sorti* is the preferred form, and samplings show it clearly dominating over *sorsi*. We get the following evidence from the standard scholarly dictionaries of the language:

- *Nykysuomen Sanakirja* lists *sorta-* uniquely as its own morphological type, which shows that *sorti* is clearly exceptional from a rule perspective. The alternative form *sorsi* is listed as a minor variant. Citations in the dictionary give 10 with *sorti*, only 1 with *sorsi*.
- *Suomen Kielen Perussanakirja* gives *sorti* as the imperfect (that is, past tense), while *sorsi* is listed as rare. There are 2 citations with *sorti*, none with *sorsi*.

Occurrences of *sorsi* are found, such as in the 10 June 2001 issue of the *Helsingin Sanomat,* the major daily newspaper in Finland, which had the following heading to a brief article: "Omat naiset sorsivat Lipposta" – that is, 'Own women oppressed Lipponen'. (Lipponen was at the time the prime minister in the government.)

Now the question is: How can we explain this preference for *sorti* over *sorsi* (the historical form), especially when all rule approaches predict *sorsi* for the standard language? One possibility is that the nearest neighbor makes the right prediction, but unfortunately the nearest neighbor is *murta-* 'to break', which takes the *tV-si* outcome, namely *mursi*. In other words, the nearest neighbor predicts *sorsi*.

When we examine the analogical set for the verb *sorta-*, we discover that the main reason for *sorti* is the *o* vowel, a factor which no analyst (Finnish or otherwise) had ever come up with prior to the AM analysis of the Finnish past tense. When we consider the analogical set, we see that the nearest gang, representing verbs ending in *rtA* (where *A* is a low vowel, either *a* or *ä*), is rather minor in its ability to provide exemplars. On the other hand, there is a huge gang, but further away, that has *o* as the first stressed vowel. Every one of these verbs takes the *V-i* outcome, thus overall *sorti* is predicted about 94% of the time. (See the schematic drawing of the analogical set for *sorta-* in Skousen, Lonsdale, and Parkinson 2002:33.)

One very important result from this example of *sorta-* is that we must not assume in advance which variables are the crucial ones and thus ignore all the others. The *o* vowel is definitely not a significant variable in the historical development of the Finnish past tense. And for all other verbs in the language, the *o* vowel is not the crucial factor. Its potential use remains latent until the appropriate given context is chosen.

7. The rule equivalent

Analogical modeling can be re-interpreted in terms of rules, as follows: (1) every possible "true" rule exists; and (2) the probability of using a "true" rule is proportional to its frequency squared. By a "true" rule, I mean that the context for the rule is homogeneous in behavior. Homogeneity occurs under two conditions: either the rule is deterministic in behavior (has only one outcome); or if the rule is non-deterministic, no subrule of the rule can behave differently.

Despite this equivalence, AM is not like regular rule approaches. First of all, there is no partitioning of the contextual space. Since all (!) the "true" rules are said to exist, there will be overlapping rules, redundant rules, and rules based on as little as one occurrence. Secondly, these equivalent "true" rules, when considered from the perspective of AM, are created "on the fly"; they are not stored somewhere, waiting to be used.

A third, and most crucial, difference is how AM treats non-deterministic rules – that is, rules that are probabilistic in nature (sometimes referred to as "variable rules"). Consider a hypothetical probabilistic rule for the past tense of the English verb *dive*. Let us assume that under some specific conditions, the probability of producing the irregular *dove* as the past tense form is 2/3, while the probability of the regular *dived* is 1/3. One immediate problem that arises in dealing with probabilistic

rules is whether the context for this rule will ever be homogeneous. It seems very unreasonable to assume that every subcontext of this probabilistic context will have precisely the same probability distribution of (2/3,1/3). In fact, it also seems very doubtful that the probability would be a nice clean ratio like (2/3,1/3). Real coins, for instance, do not have a precisely equal probability of (1/2,1/2) for coming up heads and tails. Rather, actual occurrences suggest that the ultimate objective probability for a coin coming up heads is some irrational number near 1/2, not 1/2 itself. Perhaps after many trials on a particular coin, this probability can be estimated by a rational number (say something like 0.498157). Yet even this estimate is unstable since the coin itself would be affected by all that flipping and thus the actual probability of heads would change as the coin continued to be physically tested.

The most serious issue with probabilistic rules is how to actually learn a probability and then use that probability to predict behavior. In the case of *dove/dived*, the learner would be confronted with a finite sequence of past-tense forms:

o o d d o d d o o o d o o o d d o o d o o o d o d o d o o d d d o d

(Here *o* stands for *dove* and *d* for *dived*.) From such finite sequences, a probability must be determined. One could use the sample ratio, which in this case would be (19/34,15/34). In this instance, I used a random number table to generate this sequence, and although my linguistic rule may have an objective probability of (2/3,1/3), its sample ratio is hardly ever equal to that probability. The task of determining the objective probability is very hard indeed. But the really hard part is using an objective probability, even if it is the sample ratio, to predict probabilistic behavior randomly. The random sequences used in computer programs are based on mathematical functions that look random enough over finite intervals, but ultimately involve non-randomness.

It was this major problem of learning and using probabilities that led me in 1979 to realize that the solution was much different. Instead of dealing with probabilities, it was much simpler to directly store the examples – say of *dove* versus *dived* – and then, when there was a need to predict the past tense of *dive*, to randomly select one of those previous examples, examine whether it was *dove* or *dived*, and then base the prediction on that single randomly-selected example. Thus the original motivation for analogical modeling was to solve the problem of probabilistic rules by refusing to posit them! Instead, we store and use examples. In this case, about two thirds of the examples will be *dove*, about one third *dived*. In fact, there will be no objective probabilities at all, only the occurrences.

Royal Skousen

8. Quantum computing of analogical modeling
8.1 The exponential explosion
One well-known property of analogical modeling is the exponential explosion: whenever we add one variable to the system, we basically double the running time and the memory requirements needed to determine the analogical set for a given context. This exponentiality results in pragmatic difficulties for AM since no matter how powerful the computer being used to determine the analogical set, there is a restriction on the number of variables that can be used for any one problem. The original AM program (see Skousen 1989) was restricted to only 12 variables. Subsequent work on the algorithm now allows us to posit up to about 30 variables, but there is still (and will always be) a fairly low bound on the number of variables that can be implemented on a classical computer. Various pragmatic approaches to resolving this problem have resulted in some reduction in the exponentiality, but none have been able to eliminate it.

Within the last few years, however, I have discovered striking parallels between AM and quantum mechanics, in particular quantum computing, which suggest that AM can be naturally interpreted as a variant of quantum processing. Quantum computing allows the simultaneous processing of 2^n states instead of a single state (as in a standard sequential computer). I will outline here some of these parallels.

8.2 Superpositioning, interference, and reversibility
Quantum mechanics permits n quantum bits (called qubits) to simultaneously represent 2^n states and to evolve through time by means of reversible operators. This reversibility prevents any loss of information while the 2^n states exist in this superposition of states. Such a superposition permits the multiple existence of the same "object". As the system evolves, objects can interfere and can become entangled. Further, the chances that a given state will be ultimately observed is increased or decreased. Objects that behave alike (or are "in phase") constructively interfere, while objects that behave differently (or are "out of phase") destructively interfere.

Analogical modeling follows the same basic process. Given n variables for a given context, 2^n supracontexts are defined, one supracontext for every possible combination of the n variables. As data occurrences are read in and simultaneously assigned to all the applicable supracontexts, the homogeneity of each supracontext is determined by means of reversible operators. The condition of supracontextual homogeneity prevents any increase in disagreement, thus no loss of information occurs. The same data occurrence (or "object") is typically found in more than one homogeneous supracontext, with the result that homogeneity leads to proximity and gang effects. In contrast to the constructive interference of homogeneity, all heterogeneous supracontexts are zeroed out (a kind of destructive interference).

8.3 Observation and the squaring of the underlying linearity
In quantum mechanics, the probability of each state being observed is not directly represented. Instead, a linearly defined amplitude (usually a complex number) is assigned to each state. The superposition of states and its parallel processing continues until observation (or measurement) occurs. Then the superposition collapses into a single state and the probability of that state occurring is equal to its amplitude squared.

In analogical modeling, each homogeneous supracontext is linearly represented by the number of data occurrences assigned to that supracontext. (Heterogeneous supracontexts are assigned an amplitude of zero, which means that none of their data occurrences can be accessed.) In a homogeneous supracontext, pointers are assigned between each pair of occurrences. Data input continues until the decision is made to observe the analogical set, yet when that occurs, only one of the homogeneous supracontexts is accessed. Moreover, we select a pointer to an occurrence, not an occurrence itself, with the result that the underlying linear measure based on the number of occurrences is squared.

8.4 Exemplar-based quantum computing
One interesting aspect of quantum computing is that reversibility requires that input data can never be erased, from which we may conclude that quantum computing itself is an exemplar-based system and that quantum computation of any language-based system will be an exemplar-based one. Further, analogical modeling (using random selection) appears to be a general quantum computing algorithm. The exponential explosion in AM is not inherent, but instead is the result of using classical computation.

References

Skousen, Royal. 1989. *Analogical Modeling of Language.* Dordrecht: Kluwer.
Skousen, Royal. 1992. *Analogy and Structure.* Dordrecht: Kluwer.
Skousen, Royal, Deryle Lonsdale, and Dilworth S. Parkinson. 2002. *Analogical Modeling: An exemplar-based approach to language.* Amsterdam: Benjamins.

Department of Linguistics and English Language
Brigham Young University
Provo, Utah 84602

royal_skousen@byu.edu